

Task-Oriented Complex Ontology Alignment: Two Alignment Evaluation Sets

Elodie Thiéblin Ollivier Haemmerlé Nathalie Hernandez Cássia Trojahn

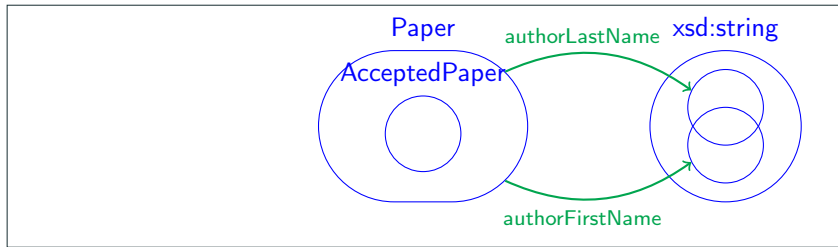
15th Extended Semantic Web Conference

IRIT & Université de Toulouse 2, Toulouse, France
firstname.lastname@irit.fr

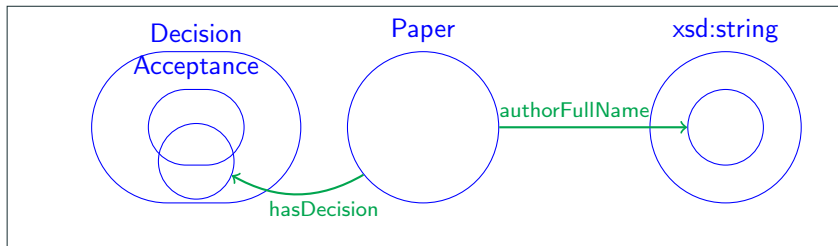
Context & Definitions

Motivating example

O_1



O_2



Retrieving all the accepted papers

o_1 $o_1:myPaper1 : o_1:AcceptedPaper$

o_2 $(o_2:myPaper2, o_2:myAcceptance) : o_2:hasDecision$
 $o_2:myAcceptance : o_2:Acceptance$

Retrieving all the accepted papers

o_1 $o_1:myPaper1 : o_1:AcceptedPaper$

o_2 $(o_2:myPaper2, o_2:myAcceptance) : o_2:hasDecision$
 $o_2:myAcceptance : o_2:Acceptance$

- Ontology merging
- Query rewriting
- Data translation

Retrieving all the accepted papers

o_1 $o_1:myPaper1 : o_1:AcceptedPaper$

o_2 $(o_2:myPaper2, o_2:myAcceptance) : o_2:hasDecision$
 $o_2:myAcceptance : o_2:Acceptance$

- Ontology merging
- Query rewriting
- Data translation

Ontology Alignments !

$$A_{o_1 \rightarrow o_2} = \{c_0, c_1, \dots, c_n\}$$

Source ontology

$$A_{o_1 \rightarrow o_2} = \{c_0, c_1, \dots, c_n\}$$

Target ontology

$$A_{o_1 \rightarrow o_2} = \{c_0, c_1, \dots, c_n\}$$

Set of correspondences

$$c_i = \langle e_{o_1}, e_{o_2}, r \rangle$$

$$c_i = \langle e_{o_1}, e_{o_2}, r \rangle$$

Relation ($\equiv, \sqsubseteq, \sqsupseteq$)

$$c_i = \langle e_{o_1}, e_{o_2}, r \rangle$$

Members : Atomic entities or constructions

Simple correspondence

- e_{o_1} atomic entity
- e_{o_2} atomic entity
- $o_1:Paper \equiv o_2:Paper$

Simple correspondence

- e_{o_1} atomic entity
- e_{o_2} atomic entity
- $o_1:Paper \equiv o_2:Paper$

Complex correspondence

- at least one of e_{o_1} , e_{o_2} contains a constructor or a transformation
- **constructor** $o_1:AcceptedPaper \equiv \exists o_2:hasDecision.o_2:Acceptance$
- **transformation function**
 $o_2:authorFullName =$
 $o_1:authorFirstName + " " +$
 $o_1:authorLastName$

Correspondences can be categorised and decomposed into **correspondence patterns** [Scharffe, 2009].

Class By Attribute Type (CAT): $A \equiv \exists b.C$

$o_1:AcceptedPaper \equiv \exists o_2:hasDecision.o_2:Acceptance$ is a CAT

Artefact for analysing the correspondences

Can guide the matching process

Ontology merging: Create a new ontology from \mathcal{o}_1 and \mathcal{o}_2 using $A_{\mathcal{o}_1 \rightarrow \mathcal{o}_2}$ and $A_{\mathcal{o}_2 \rightarrow \mathcal{o}_1}$

Query rewriting: Transformation of a query for \mathcal{o}_1 into a query for \mathcal{o}_2 using $A_{\mathcal{o}_1 \rightarrow \mathcal{o}_2}$

Ontology merging: Create a new ontology from o_1 and o_2 using $A_{o_1 \rightarrow o_2}$ and $A_{o_2 \rightarrow o_1}$

Query rewriting: Transformation of a query for o_1 into a query for o_2 using $A_{o_1 \rightarrow o_2}$

Assumption: an alignment may be different for query rewriting or ontology merging

$o_2:authorFullName = o_1:authorFirstName + " " + o_1:authorLastName$

State of the art

Complex ontology matchers

Many complex matching approaches on different kinds of schemata ¹

Here, focus on **ontology** matching

	Pattern based	No pattern
Ontology based	[Ritze et al., 2009, Ritze et al., 2010]	[Jiang et al., 2016]
Instance based	[Parundekar et al., 2010, Parundekar et al., 2012, Walshe et al., 2016]	[Qin et al., 2007, Nunes et al., 2011]

¹[http:](http://semantic-web-journal.net/content/survey-complex-ontology-matching)

[//semantic-web-journal.net/content/survey-complex-ontology-matching](http://semantic-web-journal.net/content/survey-complex-ontology-matching)

Evaluation of complex matchers so far

Approach	Dataset	Metrics
[Ritze et al., 2009]	OAEI Conference ²	Precision
[Ritze et al., 2010]		Precision
[Jiang et al., 2016]		Precision, Recall
[Parundekar et al., 2010]	DBpedia, Geonames, Geospecies, LinkedGeoData, GeneID, MGI	None
[Parundekar et al., 2012]		Precision, Recall (subset)
[Walshe et al., 2016]	DBpedia, Yago	Precision, Recall (subset)
[Nunes et al., 2011]	None	None
[Qin et al., 2007]	UMD, CMU	Precision, Recall (subset)

²<http://oaei.ontologymatching.org/>

Where we are so far

- Complex ontology matching: a developing field
- No reference dataset
- No adapted evaluation metric

Methodology and dataset

Assumptions/choices

- Equivalence preferred over subsumption
- Simple correspondences preferred over complex ones

Assumptions/choices

- Equivalence preferred over subsumption
- Simple correspondences preferred over complex ones

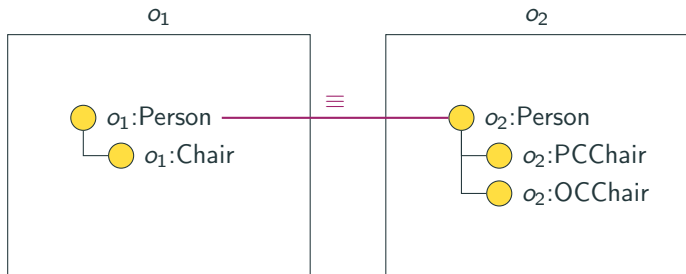
Global methodology

1. Simple equivalence correspondences $\sigma_1 - \sigma_2$
2. Creation of the complex correspondences given a task
3. Write the correspondences in a reusable format (e.g., EDOAL)

Top-down approach for ontology merging

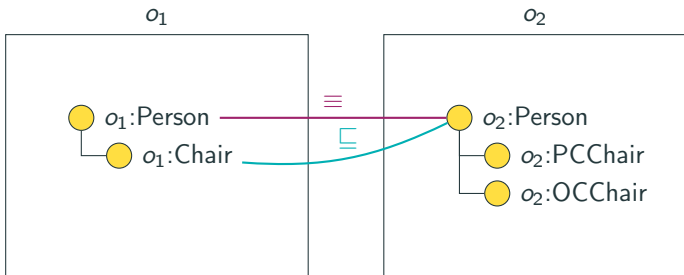
1. Create a new ontology importing o_1 and o_2
 2. Insert the simple equivalence correspondences as axioms
 3. Find for each entity of o_1 (then o_2) an equivalence or the closest super-entity not inferred
- Top-down approach : allows inference

Example of ontology merging methodology



Simple equivalence : $o_1:Person \equiv o_2:Person$

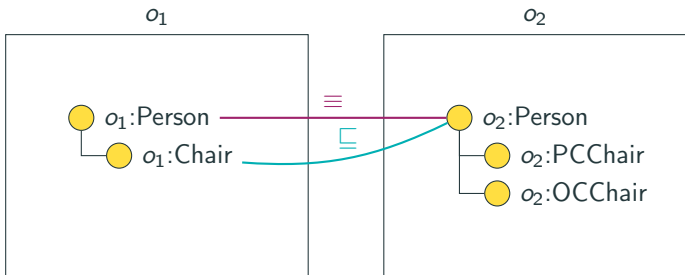
Example of ontology merging methodology



Simple equivalence : $o_1:Person \equiv o_2:Person$

Inferred axiom: $o_1:Chair \sqsubseteq o_2:Person$

Example of ontology merging methodology

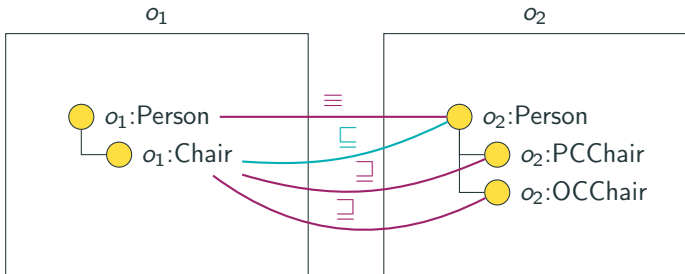


Simple equivalence : $o_1:Person \equiv o_2:Person$

Inferred axiom: $o_1:Chair \sqsubseteq o_2:Person$

- $o_1 \rightarrow o_2$: No equivalence for $o_1:Chair$ + axiom inferred

Example of ontology merging methodology



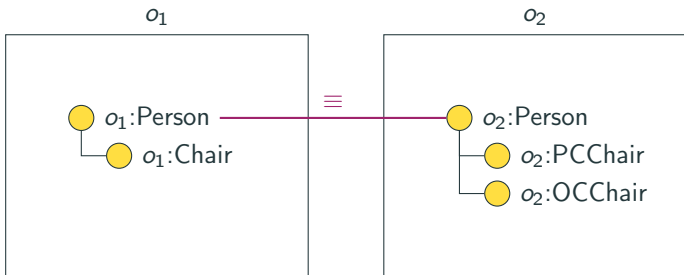
Simple equivalence : $o_1:Person \equiv o_2:Person$

Inferred axiom: $o_1:Chair \sqsubseteq o_2:Person$

- $o_1 \rightarrow o_2$: No equivalence for $o_1:Chair$ + axiom inferred
- $o_2 \rightarrow o_1$: New correspondences added
 - $o_2:PCChair \sqsubseteq o_1:Chair$
 - $o_2:OCChair \sqsubseteq o_1:Chair$

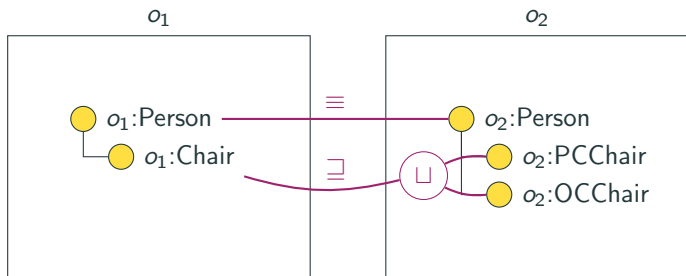
1. Find for each entity of o_1 (then o_2) an equivalence or the closest sub-entity or construction
 - The expressiveness of the alignment may prevent a reasoning task from being decidable
 - Precision favoured over recall

Example of query rewriting methodology



Simple equivalence : $o_1:Person \equiv o_2:Person$

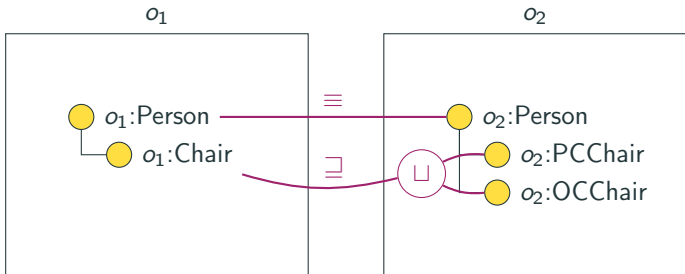
Example of query rewriting methodology



Simple equivalence : $o_1:Person \equiv o_2:Person$

- $o_1 \rightarrow o_2 : o_1:Chair \sqsupseteq o_2:PCChair \sqcup o_2:OCChair$

Example of query rewriting methodology



Simple equivalence : $o_1:Person \equiv o_2:Person$

- $o_1 \rightarrow o_2$: $o_1:Chair \sqsupseteq o_2:PCChair \sqcup o_2:OCChair$
- $o_2 \rightarrow o_1$: No new correspondences
 - $o_2:PCChair$ has no subsumed entity/construction
 - $o_2:OCChair$ has no subsumed entity/construction

Alignment sets & Evaluation

5 ontologies of the OntoFarm Conference dataset

$2 \times (10 \text{ pairs (20 oriented pairs)})$

cmt, conference, confOf, edas, ekaw

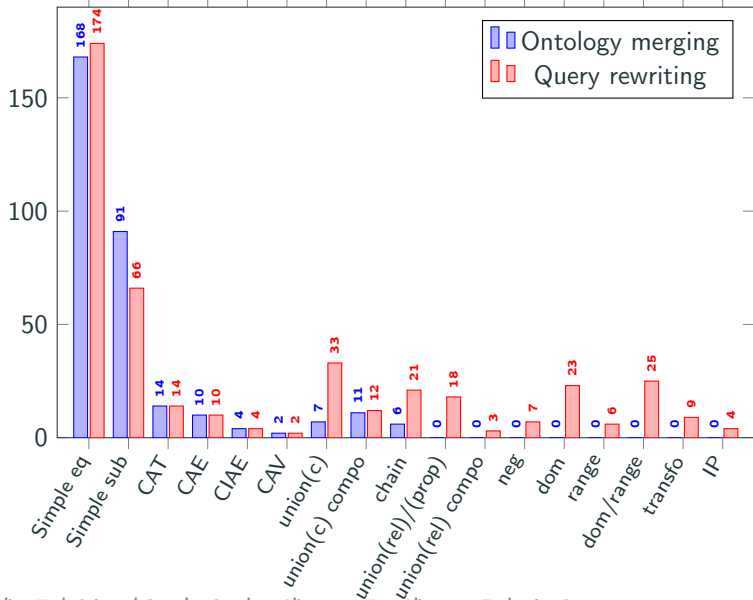
	Simple	Complex	TOTAL
Ontology merging	259	54	313
Query rewriting	240	191	431

Format: EDOAL, OWL (merging), FOL

Available under CC-BY license

<https://doi.org/10.6084/m9.figshare.4986368.v7>

Analysis of the dataset by pattern



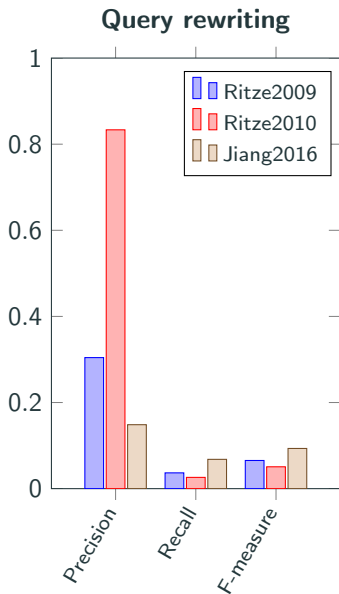
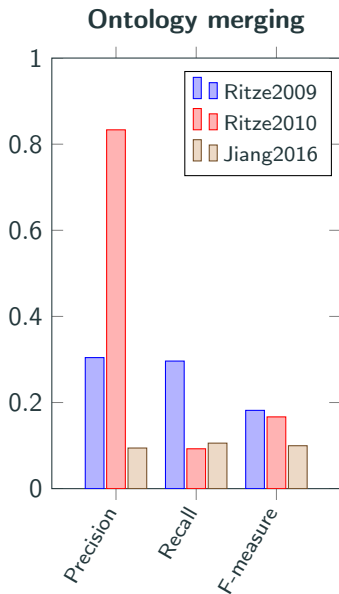
Manual evaluation of the complex correspondences (precision, recall)

Only consider the *SROIQ* correspondences for ontology merging

Consider all correspondences for query rewriting

Relation (\equiv , \sqsubseteq , \sqsupseteq) not taken into account

Manual evaluation of existing approaches on the dataset



Discussion & Conclusion

Manual creation of the alignments

Bias on the interpretation of the ontologies

Consensus dataset currently in work

Limitations (and future works)

Ontologies are not populated

No task-oriented evaluation

Metrics for automatic evaluation: semantic precision and recall ?

- Need of complex alignments
- A complex alignment evaluation dataset
- Transformation into a benchmark with instances and evaluation metrics

<https://doi.org/10.6084/m9.figshare.4986368.v7>

Thank you !

Questions ?



Euzenat, J. and Shvaiko, P. (2013).

Ontology Matching.

Springer Berlin Heidelberg.



Jiang, S., Lowd, D., Kafle, S., and Dou, D. (2016).

Ontology matching with knowledge rules.

In *Transactions on Large-Scale Data-and Knowledge-Centered Systems XXVIII*, pages 75–95. Springer.



Nunes, B. P., Mera, A., Casanova, M. A., Breitman, K. K., and Leme, L. A. P. (2011).

Complex Matching of RDF Datatype Properties.





In *6th ISWC workshop on ontology matching*.



Parundekar, R., Knoblock, C. A., and Ambite, J. L. (2010).

Linking and building ontologies of linked data.

In *ISWC*, pages 598–614. Springer.

-  Parundekar, R., Knoblock, C. A., and Ambite, J. L. (2012).
Discovering concept coverings in ontologies of linked data sources.
In *ISWC*, pages 427–443. Springer.
-  Qin, H., Dou, D., and LePendu, P. (2007).
Discovering executable semantic mappings between ontologies.
In *On the Move to Meaningful Internet Systems*, pages 832–849. Springer.
-  Ritze, D., Meilicke, C., Šváb Zamazal, O., and Stuckenschmidt, H. (2009).
A pattern-based ontology matching approach for detecting complex correspondences.
In *4th ISWC workshop on ontology matching*, pages 25–36.
-  Ritze, D., Völker, J., Meilicke, C., and Šváb Zamazal, O. (2010).
Linguistic analysis for complex ontology matching.
In *5th workshop on ontology matching*, pages 1–12.



Scharffe, F. (2009).

Correspondence Patterns Representation.

PhD thesis, Faculty of Mathematics, Computer Science and University of Innsbruck.



Walshe, B., Brennan, R., and O'Sullivan, D. (2016).

Bayes-recce: A bayesian model for detecting restriction class correspondences in linked open data knowledge bases.

International Journal on Semantic Web and Information Systems (IJSWIS), 12(2):25–52.